09550 FEB 23 1994 2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
15 June 93

3. REPORT TYPE AND DATES COVERED
Annual Progress Report 5/15/92-5/14/93

4. TITLE AND SUBTITLE
Pattern Analysis Based Models of Masking by Spatially Separated Sound Sources

5. FUNDING NUMBERS
G - AFOSR 91-0289
61102F
2313
CS

6. AUTHOR(S)
Robert H. Gilkey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wright State University
Department of Psychology
Dayton, Ohio 45435

8. PERFORMING ORGANIZATION REPORT NUMBER
AEOSR-TR- 94 0137

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
Bolling Air Force Base
District of Columbia 20332

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Research is described in three areas: masked detection, sound localization, and neural network models of sound localization. Work on masked detection indicates that substantial reductions in masking of 8 to 18 dB can be realized when the signal is spatially separated from the masker in the free-field. This reduction in masking appears to be mediated by high-frequency information. Headphone-based studies of reproducible noise masking question traditional models of binaural masking, by showing unexpected relations between responses under monaural and binaural conditions. A new response technique has been developed to support work on sound localization. Neural network models of sound localization based on binaural stimulus cues can produce responses comparable to those of human observers. Our efforts in laboratory development and in planning the Conference on Binaural and Spatial Hearing are also briefly described.

14. SUBJECT TERMS

15. NUMBER OF PAGES
14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unclassified

Pattern-Analysis Based Models of Masking by Spatially Separated Sound Sources
AFOSR 91-0289
Annual Progress Report
May 15, 1992 to May 14, 1993

I. RESEARCH OBJECTIVES

The long-term goal of this program of research is to specify the mechanisms that underlie the spatial hearing abilities of humans. One of our current projects is concerned with spatial hearing performance in the presence of noise. The motivation for this research is two fold. First, we are answering a series of basic science questions concerned with the mechanisms that allow us to "hear out" and process one particular stimulus, in the presence of other interfering stimuli. Second, the results relate to a series of applied questions, concerning the effectiveness of three-dimensional virtual auditory displays, when those displays are complex (e.g., containing many stimuli) or when they are used in a noisy environment (e.g., a cockpit). A second project is developing a model of spatial hearing. A number of potential monaural and binaural cues have been suggested as a potential basis for sound localization. We view the observer in a sound localization task as attempting to associate the pattern of acoustic cues received on a particular trial with a particular source direction. We are using neural network models to perform this pattern recognition task, and are attempting to determine which cues are necessary to achieve human-like performance.

II. STATUS OF THE RESEARCH

Much of the work described here is being conducted in the Auditory Localization Facility of the Armstrong Laboratory at Wright-Patterson Air Force Base. This facility contains a 14-foot diameter geodesic sphere, with 277 speakers mounted on its surface. This is a unique facility that allows the experimenter considerable control over the spatial distribution of sound sources when conducting sound localization or free-field masking research. Additional studies are being performed in the Signal Detection Laboratory of the Department of Psychology at Wright State University. This is a more traditional psychoacoustic facility, where subjects listen to sounds presented over headphones in individual sound-attenuating booths. Some of the work described here receives additional support from Armstrong Laboratory, from a grant from the National Institutes of Health, and through cost-sharing funds from Wright State University.

A. Detection in Noise.

Free-field Masking. Our work on free-field masking replicates previous work that has shown a substantial increase in detectability when the signal and masker are spatially separated [e.g., K. Saberi, L. Dostal, T. Sadralodabai, V. Bull, and D.R. Perrott, J. Acoust. Soc. Am. 90, 1355-1370 (1991)]. However, in our work the stimulus frequency was systematically manipulated. In these studies, the detectability of a brief click-train signal in the presence of a white Gaussian noise masker was measured as a function of the spatial separation between the signal and the masker. Both the signal and the masker were band-limited to lie within low- (below 1.4 KHz), mid- (1.2 to 6.8 KHz), or high- (above 3.5 KHz) frequency ranges. The masker was located directly in front of the subject, directly to the left of the subject, or directly above the subject.

When the signal was separated from the masker in azimuth within the horizontal plane, the detectability of the signal could be increased by as much as 18 dB (see Figure 1). Increases in detectability of as much as 8 dB were observed for separations in elevation within the median plane (see Figure 2). In all cases, the increases in detectability observed for the high-frequency signal were as great, or greater, than those observed for the low-frequency signal. Traditional models of binaural masking, based on interaural differences, did not predict the increases in detectability observed with vertical separations within the median plane, where interaural differences are relatively small. Moreover, these models seemed inadequate to explain the effects of stimulus frequency; the increase in the magnitude of the interaural level difference with increasing frequency was not great enough to predict the observed improvement in performance.

In a further attempt to relate these results to traditional headphone-based binaural masking results, continuous and gated maskers were compared. Based on the headphone results of McFadden [J. Acoust. Soc. Am. 40, 1414-1419 (1966)], who found that the "binaural release from masking" was greater with a continuous masker than with a gated masker, it might be expected that the effects of spatial separation would be greater with a continuous masker than with a gated masker. On average, we found that the signal was about 2-3 dB more detectable in the presence of a continuous masker. However, in conflict with the headphone results, these effects were not systematically related to spatial separation.

Overall, the results of these studies indicate that masking release on the order of 8 to 18 dB can be observed in free-field masking situations when the signal and the masker are spatially separated. The pattern of results from these experiments emphasizes the importance of high-frequency information. These results have important implications for display designers, indicating the nature and magnitude of the changes in detectability that can be expected when sounds are spatially separated. Portions of this work were presented at the Boston University Binaural Conference, December 1991; the fall meeting of the Acoustical Society of America in 1992; the meeting of the Human Factors Society, October 1992; and the AFOSR Review: Research on Hearing, June 1993. A proceedings paper describing some of this work has been published: Good and Gilkey (1992). A paper describing this work has been submitted: Gilkey and Good (submitted December, 1993).

<u>Reproducible Noise Masking</u>. We have also been using headphone-based masking experiments to test the predictions of traditional models of binaural interaction. The large masking level difference (MLD) observed between monaural and binaural tone-in-noise masking tasks has been used to suggest that quite different processing is employed under these two conditions (e.g., energy detection vs. interaural time processing). However, when Gilkey, Robinson, and Hanna [J. Acoust. Soc. Am. 78, 1207-1219 (1985)] examined the trial-by-trial responses of subjects, they found that the responses under the N0S0 (both noise and signal presented diotically) and N0Sπ (noise diotic, but signal presented 180° out of phase interaurally) conditions were highly correlated. That is, although the signal level under the N0Sπ condition is 10 to 15 dB lower than under the N0S0 condition (because of the MLD), individual reproducible noise-alone or signal-plus-noise waveforms that were likely to elicit a positive response (i.e., a report of signal present) under one condition were also likely to elicit a positive response under the other condition. Gilkey et al. used wideband reproducible noise samples as maskers. When Isabelle and Colburn [J. Acoust. Soc. Am. 89, 352-359 (1991)] examined the responses of subjects to narrowband reproducible noise samples, they found correlations that were much weaker and often negative. They attributed the differences between their data and those of Gilkey et al. to the differences in the bandwidth of the masker. However, Gilkey [Paper presented at the Midwinter Meeting of the Association for Research in Otolaryngology, February, 1990] directly compared narrowband and

wideband results for the same subjects and found highly significant correlations between N0S0 and N0Sπ conditions with both wideband and narrowband maskers. The correlation between N0S0 and N0Sπ responses has significant implications for models of both monaural and binaural performance. Lateralization-based models of binaural masking are unable to predict these data. On the other hand, Gilkey et al. showed that for a model such as the Equalization-Cancellation (EC) model [N.I. Durlach, "Binaural signal detection: Equalization and Cancellation Theory," in Foundations of Modern Auditory Theory II, edited by J.V. Tobias (Academic Press, New York), 371-462 (1972)] the effective maskers under the N0S0 and N0Sπ conditions are highly correlated. Thus, the observed correlation of responses between these conditions is not necessarily unexpected.

In order to evaluate more fully the predictions of the EC model, we have measured the responses of subjects to individual reproducible waveforms under conditions where the effective masker at the output of the EC device should be quite different from the masker under the N0S0 condition. Under the NuSπ condition (independent noises to the two ears, signal 180° out of phasc interaurally), the EC device should subtract the stimuli arriving from the two ears, such that the effective masker at the output of the EC device is the difference between the two monaural maskers. Thus, the EC model predicts that N0S0 responses to either of the two "monaural" maskers should be only partially correlated with the NuSπ responses. This is exactly what we observed in the data of our subjects. However, when the responses to the two monaural waveforms were averaged and then compared to the NuSπ responses, a very high correlation was observed. This result was not anticipated.

Because we were surprised by this result, we next examined a condition where the actual masker under the N0S0 condition was the difference between the two maskers presented under the NuSπ condition. That is, the masker under the N0S0 condition was the predicted effective masker under the NuSπ condition. Rather than the strong correlation we expected between these two conditions, based on the EC model, only a weak relation was observed.

Overall, the pattern of results suggested two possibilities, either: 1) the NuSπ condition is not a true binaural condition, as suggested by Durlach, Gabriel, Colburn, and Trahiotis [J. Acoust. Soc. Am. 79, 1548-1557 (1986)], or 2) the EC model is an inadequate model of binaural hearing. Parts of this work were presented at the Boston University Binaural Conference, December, 1991, and at the Midwinter Meeting of the Association for Research in Otolaryngology, February, 1993.

B. Localization in Noise.

In many situations the observer must be able to determine the direction of the sound source, once it has been detected. Presumably, because of the increased complexity of the localization task relative to the detection task, a more complete representation of the signal information would be necessary for accurate localization. Much of our work on sound localization is represented by the thesis research of Michael D. Good. He is investigating localization accuracy in noise as a function of signal-to-noise ratio and as a function of the location of the masker.

In designing these experiments, we realized that currently available response techniques allow responses to be collected at only very slow rates (2-4 responses per minute). It was clear that if we were to collect data at these slow rates, even relatively simple experiments would take months or years to complete. Therefore, we needed to develop a technique that would allow subjects to record accurately the perceived location of a sound, at speeds that were much more rapid than was possible with current techniques.

After considering a number of alternatives, we decided that a pointing technique would be the most effective procedure. It was known that subjects could accurately indicate the location of a sound by verbally reporting spherical coordinates [F.L. Wightman and D. Kistler, J. Acoust. Soc. Am. 85, 868-878 (1989)]. Our own pilot studies indicated that when presented with spherical coordinates, subjects could point to the corresponding location on a small plastic hemisphere to within a few degrees. We therefore developed a technique in which the subjects indicate the perceived location of a sound by pointing at an 8-in spherical model of auditory space. The subjects point at the sphere using a magnetic stylus, whose XYZ coordinates are monitored with a Polhemus Fastrack "head tracker"; they then press a foot-switch to record their response. The results indicate that subjects are able to respond at rates of 16-19 responses per minute, considerably faster than with other techniques. Further, the accuracy of their responses is comparable to that which Wightman and Kistler observed with the verbal reporting technique. Figures 3 and 4 compare the azimuth and elevation judgment centroids of our subjects to the judgment centroids of two of the subjects of Wightman and Kistler. These results were presented at the AFOSR Review: Research on Hearing, June 1993. A paper describing this technique has been submitted [Gilkey, Good, Ericson, Brinkman, and Stewart, (submitted)].

C. Neural Network Models of Sound Localization.

At least three sources of acoustic information are generally recognized as providing the foundation for sound localization: interaural time differences, interaural level differences, and direction-specific spectral modulations introduced by the acoustics of the torso, head, and pinnae. No model has been developed to describe how these disparate sources of information are combined into a single unified perception of the source location.

If the pattern of interaural time differences, interaural level differences, and spectral modulations is unique for each source direction, then the task of the observer in a localization experiment can be viewed as estimating the value of these cues and determining the location that corresponds to the estimated pattern; that is, within this view, sound localization is a pattern recognition task. Because neural networks have had great success in solving other pattern recognition problems, we have been using them to model sound localization.

In our initial investigations, the model has been composed of a preprocessing section and a neural network section. In the preprocessing stage, the click signals were convolved with head-related transfer functions (filters that simulate the acoustic effect of the torso, head, and pinnae). The filtered clicks were then corrupted by internal noise; each point on the waveform was multiplied by a random amplitude jitter and subjected to a random delay, in a manner similar to that described by Durlach [J. Acoust. Soc. Am. 35, 1206-1218 (1963)]. A broadband cross-correlation was computed between the jittered waveforms in the left and right ears and the lag corresponding to the maximum in the cross-correlation function was one input to the network section of the model. In addition, Fast Fourier Transforms were computed from the waveforms in the left and right channels and the logarithm of the energy in each of 22 rectangular quarter-octave bands was determined. The difference between the log spectra in the left and right ears provided 22 additional inputs to the network stage. The sound source could originate from any of 144 directions ranging in azimuth from -165° to +180° and in elevation from -36° to +54°. One hundred training vectors were generated for each of the 144 source locations (a total of 14,400 training vectors). A second set of 14,400 vectors was used as a test set. There were 23 input units, 50 hidden units, and 30 output units. A fully connected feed-forward network was trained, with back-propagation, to "turn on" 1 of 6 output units to indicate which of 6 possible elevations had been presented, and 1 of 24 output units to indicate which of the 24 possible azimuths had been presented.

Figure 5 shows a comparison of the responses of a human subject and of the model in comparable listening conditions. The top two panels show the azimuth component of the judgment centroid plotted as a function of the target azimuth. As can be seen, both the network model and the human subject made very accurate azimuth judgments. The bottom two panels show the elevation component of the judgment centroids as a function of the actual elevation. The overall performance of the human and the model are similar, but the human systematically overestimates the elevation, while the model underestimates high elevations and overestimates low elevations. (This results, in part, from the response restrictions placed on the model; that is, it cannot respond with elevations greater than 54° or less than 36°; the human was not similarly constrained). The average angle of error for the human and the model are similar in magnitude and the number of front/back reversals observed for the model and the human are also comparable.

Wightman and Kistler [J. Acoust. Soc. Am. 91, 1648-1661 (1992)] demonstrated that for human observers low-frequency timing information plays a dominant role in determining their localization judgments. That is, when interaural time cues provide information about the source location that conflicts with information provided by interaural level differences or "spectral cues," subjects tend to judge the sound as coming from the location indicated by the interaural time difference, rather than from the location indicated by these other cues. Following Wightman and Kistler, we took the previously described network (trained on "normal" stimuli) and tested it on stimuli with phase spectra that had been modified to correspond to the phase spectrum of a sound coming from 0° azimuth and 0° elevation, from -45° azimuth and 0° elevation, or from 90° azimuth and 0° elevation. The pattern of errors observed for the model was quite similar to that observed for Wightman and Kistler's human subjects. That is, in most cases, the model responded with the location indicated by the phase spectra, rather than the location indicated by the power spectra.

In this modeling effort, we use the neural network in a role similar to that of an "ideal detector"; thus, the implications of this work are not in terms of the structure of the neural network itself. Rather we are determining the viability of various acoustic cues for mediating human-like performance. Thus far, this work indicates that for the simple stimuli employed here, there is sufficient information in the binaural cues alone to produce localization performance comparable to that of humans. Portions of this work were presented at the fall meeting of the Acoustical Society of America in 1992; the Boston University Binaural Conference, December 1992; and the AFOSR Review: Research on Hearing, June 1993.

D. Laboratory Development.

Armstrong Laboratory. In support of our localization research, the new pointing response technique, described previously, was developed. Experiment control software was developed, which incorporates this technique, along with the ability to present sounds from random directions selected from a set of up to 239 speakers. In addition, a second source (i.e., a masker) can be presented from any of 5 fixed locations.

Wright State University. Two DECsystem 3000 Model 400 workstations were purchased to provide data analysis, graphics, and modeling capabilities for the laboratory.

E. Conference on Binaural and Spatial Hearing

Considerable effort during this grant period was expended on planning the Conference on

Binaural and Spatial Hearing, which was held at the Hope Hotel and Conference Center at Wright-Patterson Air Force Base in Ohio, on September 9-12, 1993. Speakers at the conference included Timothy Anderson, Leslie Bernstein, Jens Blauert, John Brugge, Thomas Buell, Mahlom Burkhard, Robert Butler, Rachel Clifton, Steven Colburn, Theodore Doll, Richard Duda, Nathaniel Durlach, Raymond Dye, Mark Ericson, Scott Foster, Robert Gilkey, Wesley Grantham, Ervin Hafter, William Hartman, Janet Koehnke, Armin Kohlrausch, Birger Kollmeier, Gregory Kramer, Shigeyuki Kuwada, Richard McKinley, Donald Mershon, John Middlebrooks, David Perrott, Edgar Shaw, Kourosh Saberi, Barbara Shinn-Cunningham, Richard Stern, Elizabeth Wenzel, Frederic Wightman, Tom Yin, William Yost, and Eric Young.

III. PUBLICATION ACTIVITY

Papers Published

Good, M.D., & Gilkey, R.H. (1992). Masking between spatially separated sounds. Proceedings of the 36th Annual Meeting of the Human Factors Society, 1, 253-257.

Papers in preparation

Gilkey, R.H., & Good, M.D. Effects of frequency on free-field masking. Submitted to Human Factors, December, 1993.

Gilkey, R.H., Good, M.D., Ericson, M.A., Brinkman, J., & Stewart, J.M. A pointing technique for rapidly collecting localization responses in auditory research. Submitted to Behavior Research Methods, Instrumentation, and Computers, December, 1993.

IV. PARTICIPATING PROFESSIONALS

Robert H. Gilkey

De Anza College, Cupertino, CA			
University of California, Berkeley, CA	B.A.	1976	Psychology
Indiana University, Bloomington, IN	Ph.D.	1981	Psychology

Dissertation title: "Molecular psychophysics and models of auditory signal detectability."

V. INTERACTIONS

Conference presentations and invited talks

Gilkey, R.H. (1993). Pattern-analysis based models of masking by spatially separated sound sources. AFOSR Review: Research in Hearing, Fairborn, OH, June.

Gilkey, R.H. (1993). Comparing predictions of the Equalization-Cancellation Model to NuSπ performance in a reproducible noise masking task. Association for Research in Otolaryngology Mid-Winter Meeting, St. Petersburg Beach, FL, February.

Gilkey, R.H. (1993). Auditory space perception and virtual environments. Ohio Consortium for Virtual Environment Research, Dayton, OH, January.

Gilkey, R.H., Janko, J.A., & Anderson, T.R. (1992). Using neural nets to model sound localization. Boston University Binaural Conference, Boston, MA, December.

Gilkey, R.H., & Good, M.D. (1992). Effects of frequency and masker duration on free-field masking. Journal of the Acoustical Society of America, 92, 2334(A).

Anderson, T.R., Janko, J.A., & Gilkey, R.H. (1992) An artificial neural network model of human sound localization. Journal of the Acoustical Society of America, 92, 2298(A).

McKinley, R.L., Ericson, M., Perrott, D., Brungart, D., Gilkey, R., & Wightman, F. (1992). Minimum audible angle for synthesized localization cues presented over headphones. Journal of the Acoustical Society of America, 92, 2297(A).

Gilkey, R.H. (1992). The correlation between responses under monaural and binaural conditions. Journal of the Acoustical Society of America, 92, 2298(A).

Good, M.D., & Gilkey, R.H. (1992) Masking between spatially separated sounds. The 36th Annual Meeting of Human Factors Society, Atlanta, GA, October.

Figure captions

Figure 1. Threshold signal-to-noise ratio is plotted as a function of the azimuth of the signal within the horizontal plane. The masker was presented from directly in front of the subject, 0° azimuth, 0° elevation (left panel), or from directly to the left of the subject, -90° azimuth and 0° elevation, (right panel). Threshold estimates have been averaged across the three subjects. Negative values along the abscissa indicate positions to the left of the subject and positive values indicate positions to the right of the subject. A value of 0° indicates a speaker location directly in front of the subject. The position of the arrow shows the location of the masker.

Figure 2. Threshold signal-to-noise ratio is plotted as a function of the elevation of the signal within the median plane. The masker was presented from directly in front of the subject, 0° azimuth, 0° elevation (left panel), or from directly above the subject, 0° azimuth and 90° elevation (right panel). Threshold estimates have been averaged across the three subjects. Negative values along the abscissa indicate positions below the horizontal plane and positive values indicate positions above the horizontal plane. Elevations greater than 90° indicate locations in the rear hemisphere. The position of the arrow shows the location of the masker.

Figure 3. The azimuth coordinate of the judgment centroid for each target location is plotted as a function of the azimuth of the target. The top three panels show data for each of the 3 subjects in our experiment; they responded with the pointing technique. The bottom two panels show data for 2 of the subjects of Wightman and Kistler; they responded verbally. (The panel on the bottom left shows data from one of their better subjects and the panel on the bottom right shows data from one of their worst subjects.) The centroids in the top panels are based on 8 judgments at each speaker location. The centroids in the bottom panels are based on either 6 or 12 judgments at each speaker location. Front-back reversals have been resolved.

Figure 4. The elevation coordinate of the judgment centroid for each target location is plotted as a function of the target elevation. Note that the range of values on the axes has been reduced substantially relative to Figure 3. Other details are as in Figure 3.

Figure 5. Performance of subject SDO, from the study of Wightman and Kistler (1989), and performance of a neural network, receiving only binaural input, are compared. In the top two panels, the azimuth coordinate of the judgment centroid is plotted as a function of the target azimuth. In the bottom two panels, the elevation coordinate of the judgment centroid is plotted as a function of the target elevation. The panels on the left show the results for subject SDO. The panels on the right show the results for the neural network.

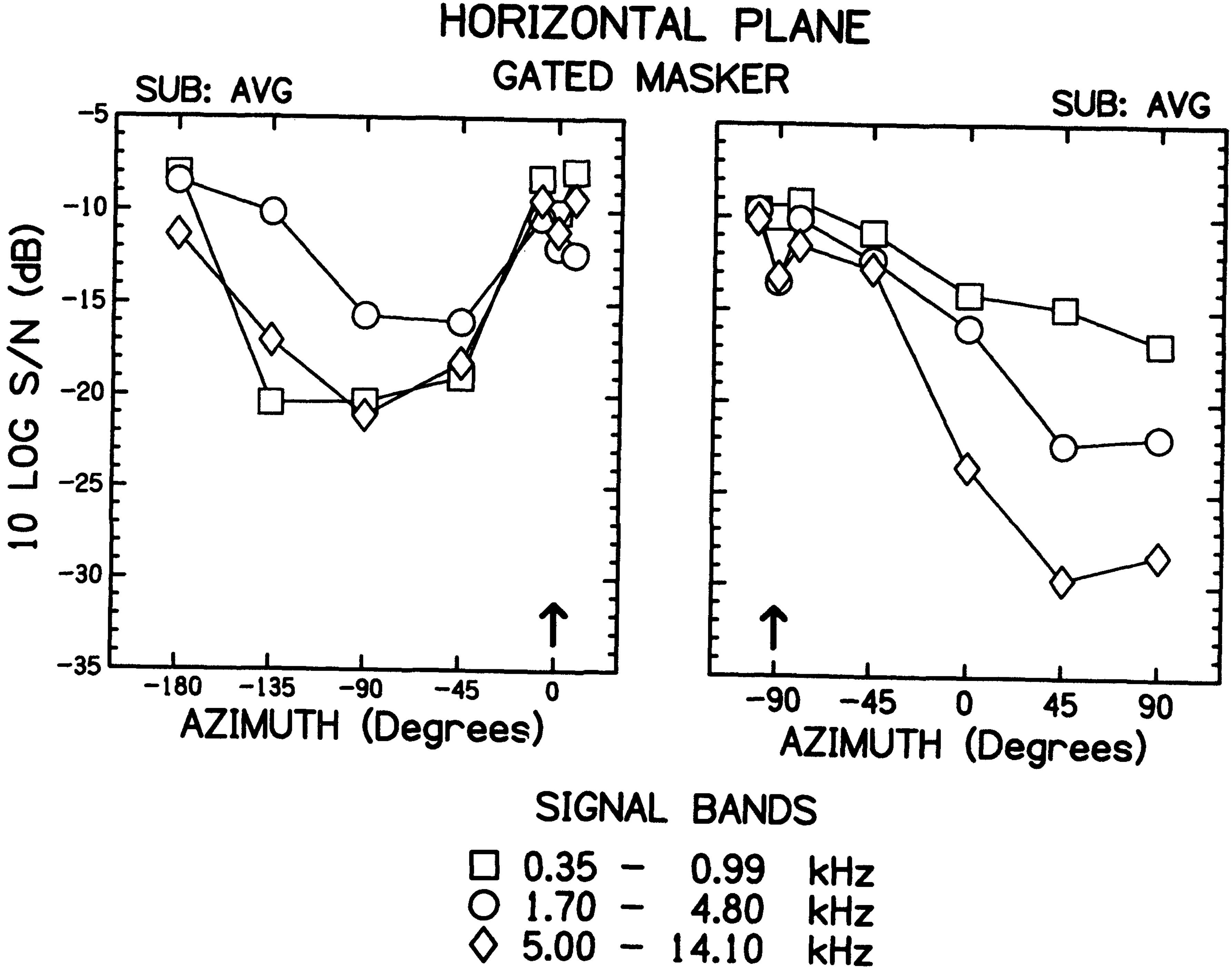

HORIZONTAL PLANE
GATED MASKER
SUB: AVG
SUB: AVG
10 LOG S/N (dB)
-5
-10
-15
-20
-25
-30
-35
-180
-135
-90
-45
0
AZIMUTH (Degrees)
-90
-45
0
45
90
AZIMUTH (Degrees)
SIGNAL BANDS
□ 0.35 - 0.99 kHz
○ 1.70 - 4.80 kHz
◇ 5.00 - 14.10 kHz

Figure 1

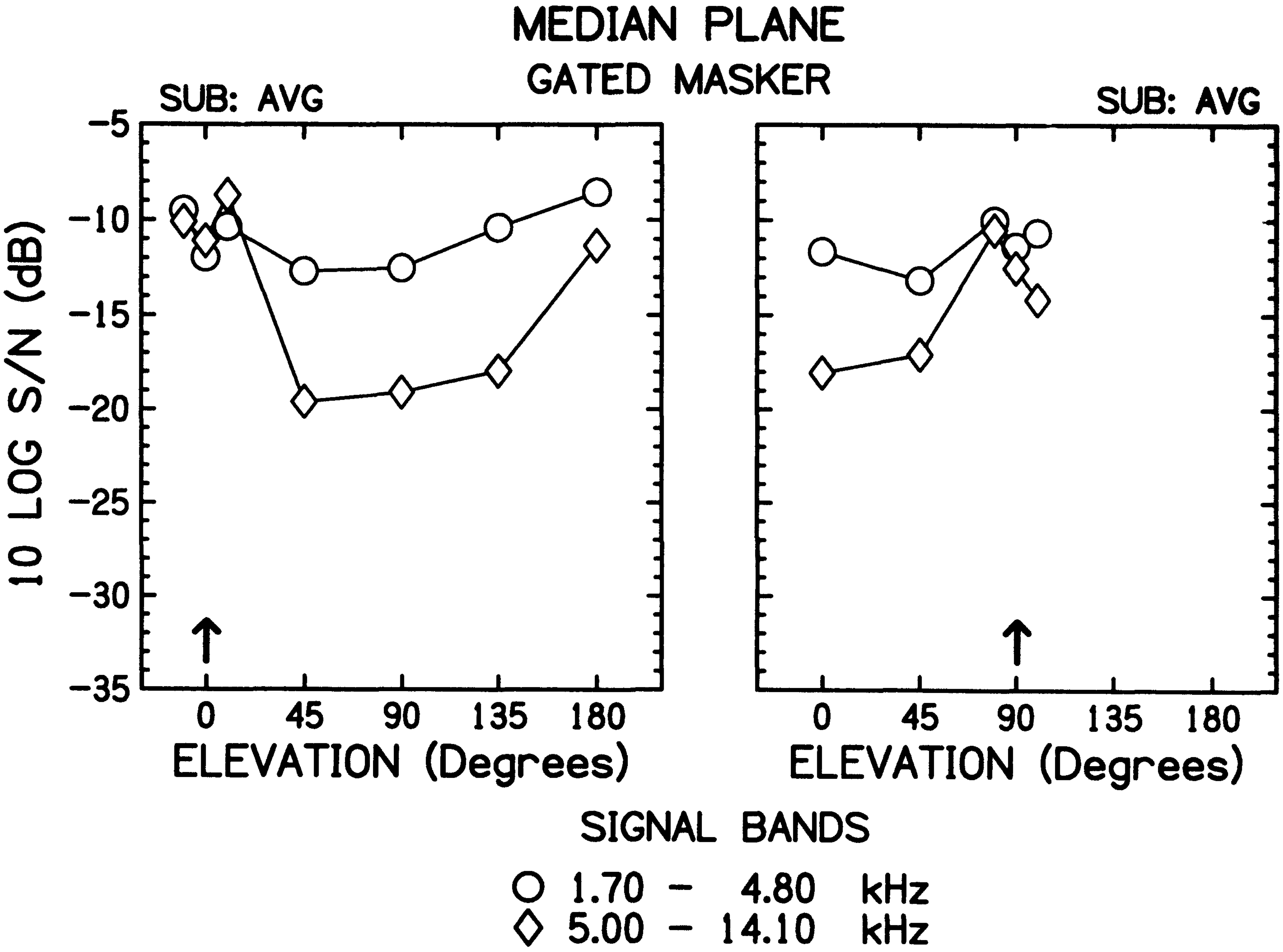

MEDIAN PLANE
GATED MASKER
SUB: AVG
SUB: AVG
10 LOG S/N (dB)
-5
-10
-15
-20
-25
-30
-35
0
45
90
135
180
ELEVATION (Degrees)
0
45
90
135
180
ELEVATION (Degrees)
SIGNAL BANDS
○ 1.70 - 4.80 kHz
◇ 5.00 - 14.10 kHz

Figure 2

Figure 3

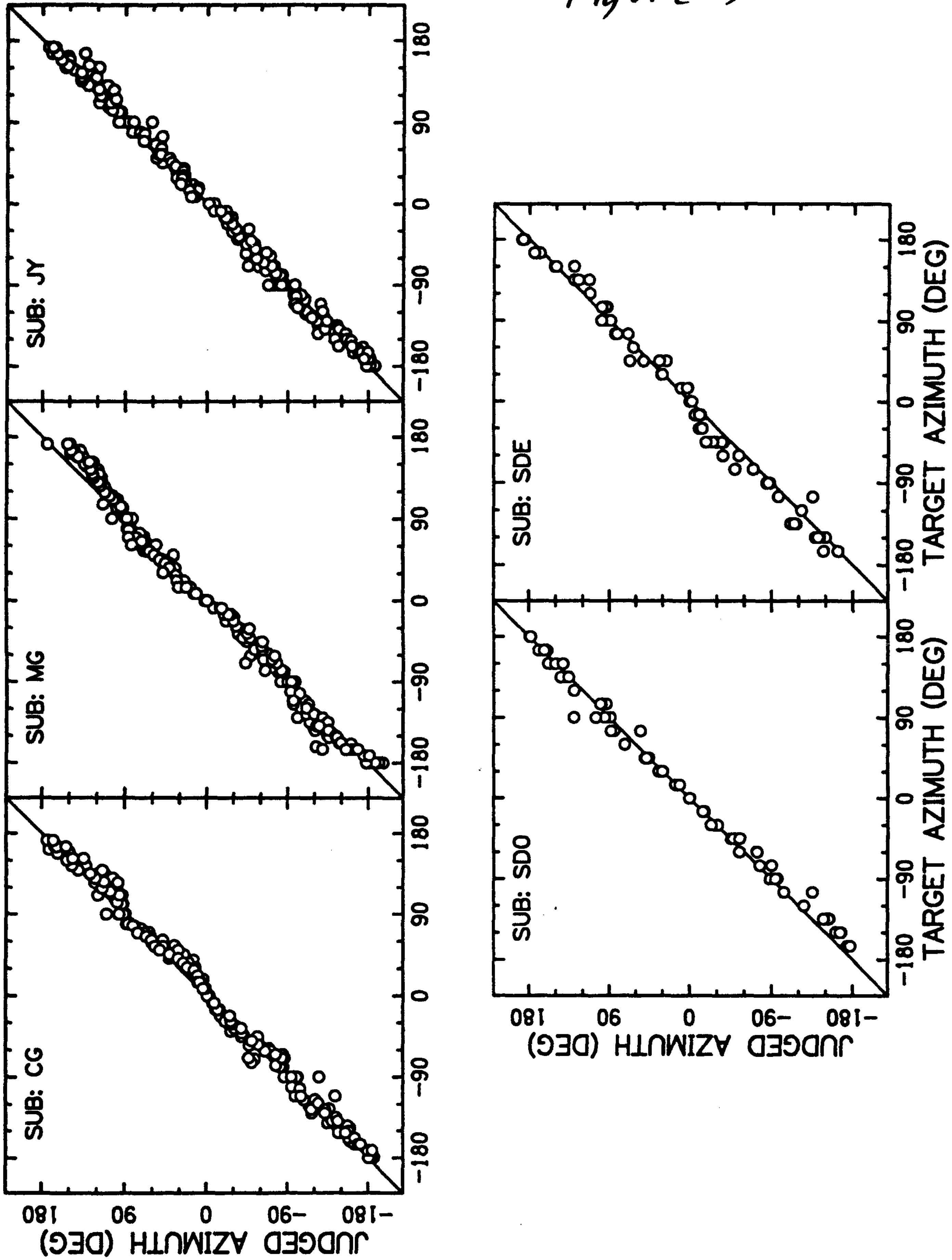

SUB: JY
SUB: MG
SUB: CG
SUB: SDE
SUB: SDO
180
90
0
-90
-180
JUDGED AZIMUTH (DEG)
TARGET AZIMUTH (DEG)

Figure 4

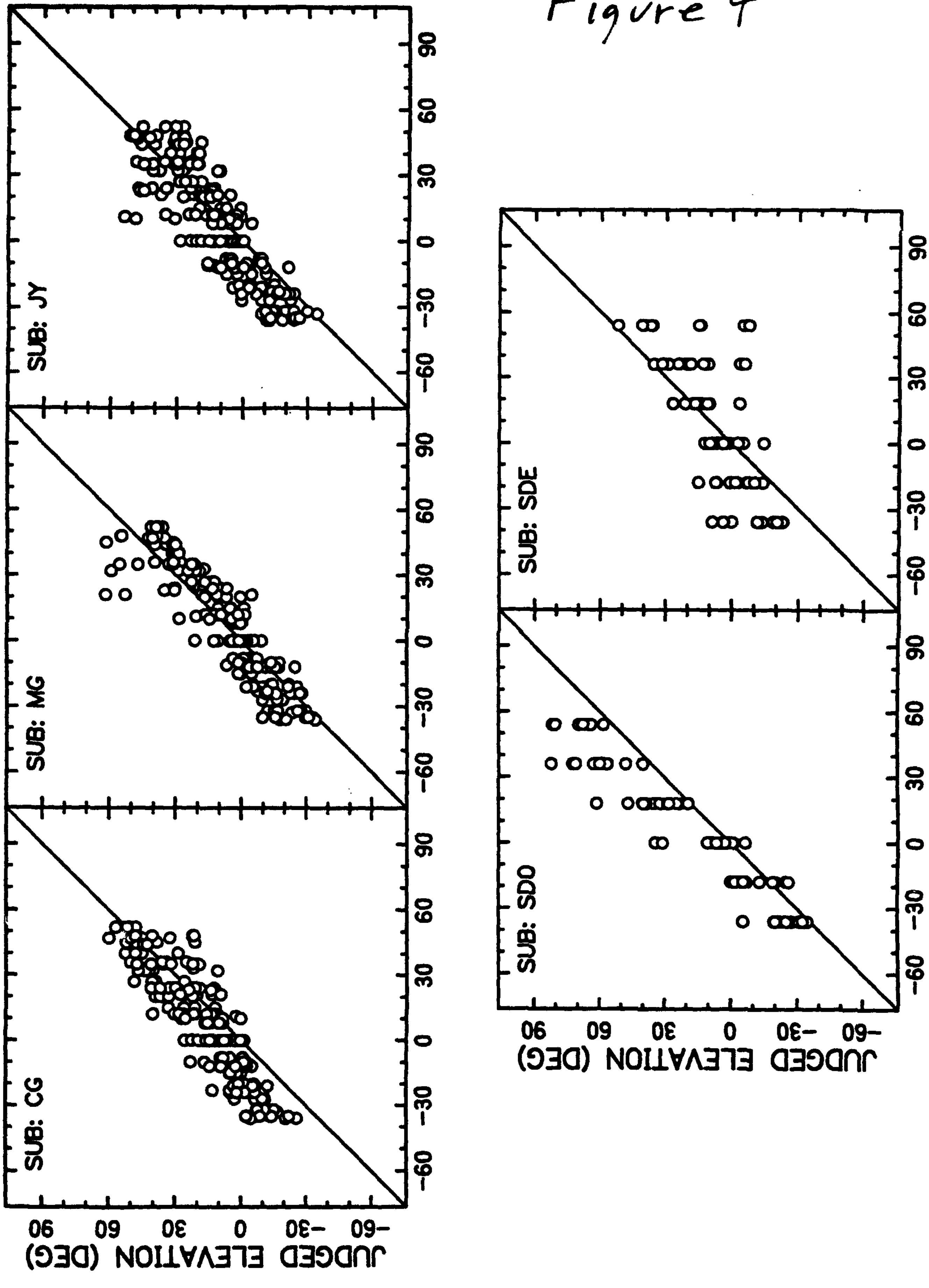
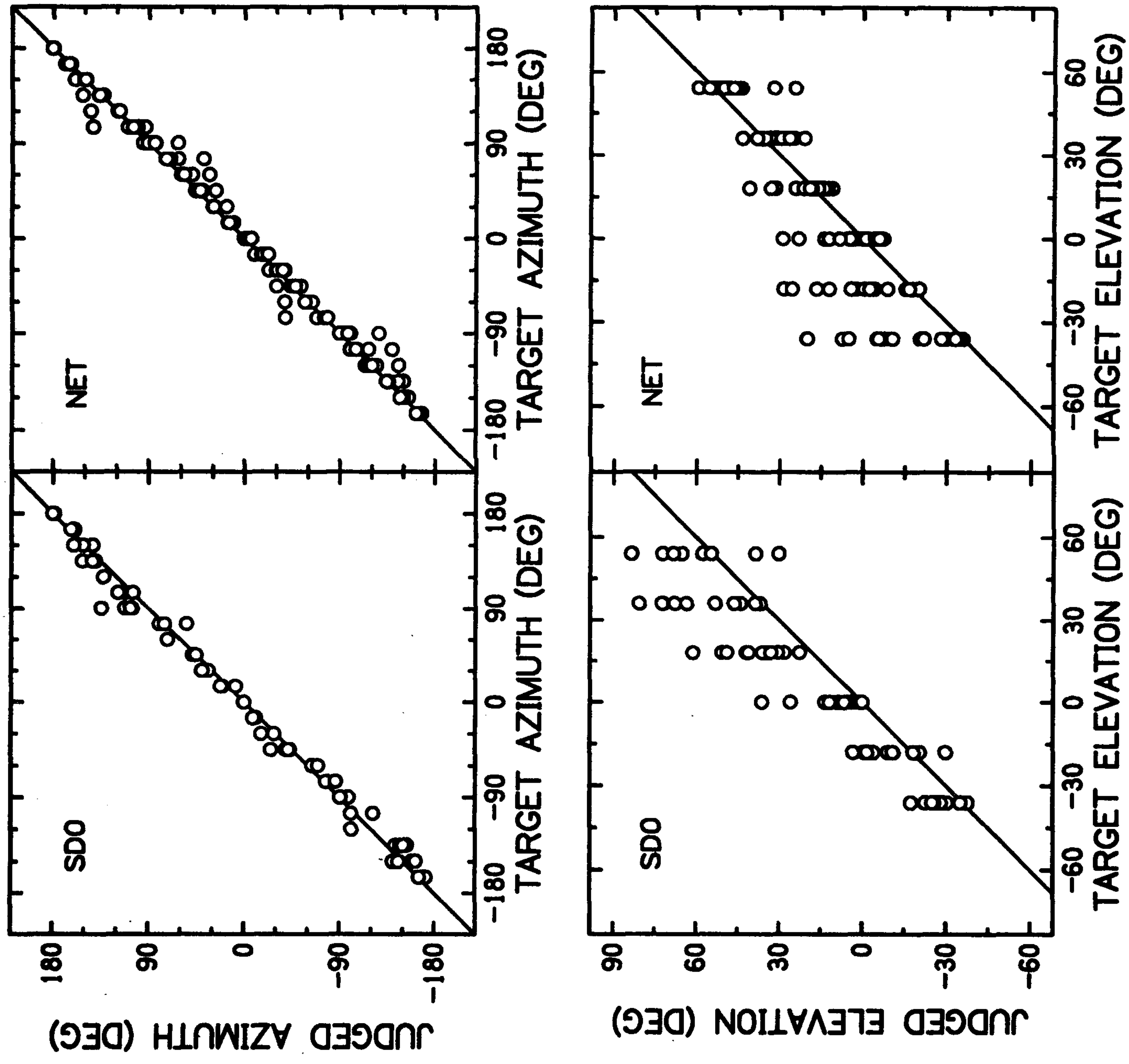

SUB: JY
SUB: MG
SUB: CG
SUB: SDE
SUB: SDO
TARGET ELEVATION (DEG)
TARGET ELEVATION (DEG)
JUDGED ELEVATION (DEG)
JUDGED ELEVATION (DEG)
-60
-30
0
30
60
90

Figure 5


SDO
NET
SDO
NET
JUDGED AZIMUTH (DEG)
180
90
0
-90
-180
TARGET AZIMUTH (DEG)
-180
-90
0
90
180
JUDGED ELEVATION (DEG)
90
60
30
0
-30
-60
TARGET ELEVATION (DEG)
-60
-30
0
30
60